COMPARISON OF HUMANS AND ROBOTS EMOTION EVALUATION PERCEPTION BASED ON HUMAN SOUND

Vesna Kirandziska

Faculty of Computer Science and Engineering (FCSE)
Skopje, Macedonia

Nevena Ackovska

Faculty of Computer Science and Engineering (FCSE)
Skopje, Macedonia

ABSTRACT

In human-to-human interaction emotion perception is the perception on the emotion of the other people, which, due to the nature of emotions is not so precise. On the other hand, perception on emotions in human-computer interaction is still an open problem. A lot of work is done in direction of finding suitable model for perceiving emotions based on different input signals and classification models. Here, only sound signals are considered. Emotion perception is being used in human-robot interaction among others. Robots are machines that can partially act like humans, and today some robots can even perceive emotions. Here a custom build emotion aware robot that perceives emotion evaluation is used to investigate the similarity and differences of the robot's perception with human's perception. This paper presents a discussion about the future of robots that can perceive human emotions and their usage.

I. INTRODUCTION

In recent years lots of robots have been designed. By one possible classification of robots there are mechanical robots, mobile robots, nanorobots, humanoid robots etc. Mobile robots can move in the environment, while humanoid robots tend to look and act (partially) like humans. Sony AIBO robot, Zeno, HPR-45 and ASIMO robot are only few examples of existing humanoid robots ([2], [13]). In general, robots can come with the looks of a pet (dog or cat), a human (cartoon character, man or woman) or part of a human body (arm, legs or head).

From a robot's usage point of view, there are industrial robots, personal robots and military robots. Personal robots can be used for entertainment, educational purposes, healthcare etc. In the future robots can be seen as part of everyday human's life. In general, being surrounded with robots, humans should interact with these machines. Consequently, the main feature of robots (especially personal robots) is their interaction with humans. The field of computer technology that considers this feature is Human-robot interaction. Human-robot interaction is manly the topic of this paper.

We are considering the human-like interaction with robots as a final goal. Human's interaction with robots should be based on natural conventions like natural language or social rules (facial expression, mimic and body movements). Robots should no longer be just indifferent logical machines, but they could become capable of understanding human's feelings, needs and desires. In this paper we are considering one of the most important information that is needed for robots to understand humans: emotions. In the other direction, the influence of emotion perception on human living is considered.

Classification of emotion is gaining attention due to the widespread applications into various domains. Examples of applications that can use emotion classification are video and computer games, human-to-robot interaction systems and even more call centers ([2], [4], [5], [13]). In human-to-robot interaction emotions are especially significant because emotion perception can facilitate communication between these two subjects.

In general, information from speech ([1], [3], [9], [12]), facial expression [14] and brain activity has been used for emotion classification so far. Here, only sound signals are considered due to the usage of emotion classifiers in call centres. Proofs that sound features posses indicators for human emotions are found in [7]. Also, the results in [6] show that certain emotions are correlated with the values of some sound features extracted from human speech, like pitch for example.

The research presented in this paper will be exposed in the following sections. The first section describes emotion in general. The next section describes how robots can perceive sound. In the following section, elaboration of robots that perceive emotions is presented. Also the robot used for this research is introduced. Afterward, we present the experimental results for comparison of human's and robot's emotion perception. Next, a discussion about the influence and effects of these robots in the environment is given. In the end, the future work with the robot and its interaction with the environment are analyzed.

II. EMOTIONS

The definition on emotion is vague. Today, many different types of emotions are recognized. Interesting question is how emotion can be represented. One important question for emotion representation is weather the emotion is a qualitative or quantitative measure.

Emotions viewed as qualitative measure can be represented as a set of emotion types. Paul Ekman [5] defined six basic emotions: anger, disgust, fear, happiness, saddens and surprise. In [7] emotions are split in ten different categories: Pleasant, Activity, Potency, Anger, Boredom, Disgust, Fear, Happiness, Sadness and Surprise. In different studies different emotions are recognizes. Today more than twenty different emotion categories are known.

The other view on emotions represents types of emotions (categorical data) as points in a multidimensional space. Plutchik used a two-dimensional space called the activation-evaluation space [5] shown on Fig. 1. Activation describes the degree of emotion expression: high or low, while evaluation describes the polarity of the emotion: positive or negative.

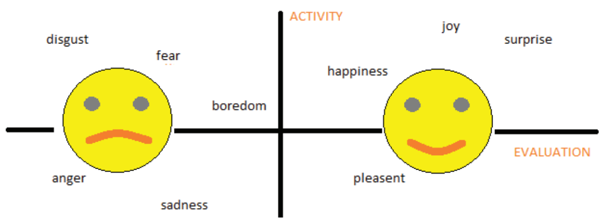

Figure 1: Emotions in activation-evaluation space.

In this research we consider only one dimension - evaluation. Emotions here are classified in two categories: positive or negative emotion. The ordinate on the graph given on Fig. 1 is the evaluation axis. Right of the origin, the positively evaluated emotions categories are represented. Accordingly, left of the origin, the negatively evaluated emotions are represented. As a result of representing emotions by the evaluation axes, in machine learning the problem of emotion perception can be viewed as a classification problem.

A different three-dimensional space, called the Lövheim cube of emotions, for representing emotions exists. Each axis represents the quantity of one of the chemical substances: Serotonin, Dopamine and Noradrenalin that are created when a human organism is in some emotional state.

There is versatile human behaviour that is common for some emotion expression among humans. As an example, in one psychological study a dependency of the happy emotion and the human pitch is discovered [6]. The results showed that when people are happy they have lower pitch.

However, because of the human’s individual and cultural differences, there could be some differences in the emotion manifestation. These can be caused by the environment. This dependency with the environment is one problem in emotion perception.

Another problem in perceiving emotions is the ambiguity [6]. Indeed, one person could have many emotions simultaneously. This makes the emotion perception harder for modelling.

Today we are not yet agreed on which sound features are most important for the emotion classification problem. Lots of features have been used in different studies ([3], [10], [11], [16]). Some important sound features for emotion classification are described in [7].

Sound features, based on their nature, can be classified as intensity features, pitch features, quality features and tempo features. The bases of all these features can be perceived by humans, and that is one reason why these are important for the classification model. Truly, in the psychological researches made by Picard a dependency of some basic features and emotions is shown [6].

III. EMOTION AWARE ROBOTS THAT PERCEIVE SOUND

A. Perceiving sound in human-robot interaction

Human-robot interaction is a two way communication. In one direction robot perceives humans’ actions and acts accordingly. In the other, robot’s behaviour is perceived from humans and is acted upon. Many signals that carry information about human’s state and behaviour are emitted by humans. These signals can be received and measured with appropriate devices. For example a camera can be used by the robot to gain a visual image of humans in the interaction. Furthermore, human sound can serve as a signal from which information about the human speech and voice in general can be perceived by robots. Many other signals exists that carry information and that can be used in many different situations.

Robots that can perceive sound should have an input device that records sound signals from the environment. These signals are indirectly signals emitted from humans. As a consequence, the sound should be accordingly preprocessed. The general goal of robots that perceive sound is to extract the most valuable sound features that are needed for the robot to process. There are many different sound features that can be calculated form sound signals. These are mostly based on the amplitude, the pitch and the tempo. There are several algorithms for calculation of different sound features. Usually for an action to be taken, a robot uses only some of those features. Indeed, this depends on the usage of the robot.

As stated previously, in this paper we are primarily concerned with the information for human emotion. Which signals are relevant for emotion perception is an important question for building robots that can recognize human emotions. New psychological theories state that emotions can be manifested by subjective experience, peripheral physiological reactions and motor expression. Humans show their inner emotions in several ways: through mimic, using movements, gestures, but through human voice as well. Indeed, speech possesses features that are indicators of human emotions [8]. According to Picard certain emotions are correlated with the values of some sound features extracted from human speech like pitch mean, amplitude standard deviation or tempo. For example, when people are happy they speak faster and have lower pitch. However, because of the human's individual and cultural differences there could be some differences in the emotion manifestation in the human speech. These are caused by the environment and genes factors. Also, according to some psychological researches, one person could have many emotions simultaneously, and as a result different people perceive his/her emotions differently. This is big problem named emotion ambiguity. Because of these and other difficulties emotion perception is still an open problem. In computer science and information computer technology the emotion perception problem can be narrowed down to a classification problem. This implies creating an automatic emotion perception system that chooses one class of emotion.

So far, information from speech, facial expression and brain activity has been used for automatic emotion perception. Although speech and sound signals are good candidates for perceiving emotional reactions, there exists not satisfactory accurate automatic emotion perception software.

B. *Perceiving emotion in human-robot interaction*

As disused in the previously, robots that perceive emotion from human speech can be built. These robots must have an appropriate hardware. In addition, their software must include automatic emotion perception system that decides on human emotions. This is not an easy task, since there are around twenty different types of emotions. These emotions can be put in the two dimension evaluation-activation space. Emotion activation shows the degree of emotion expression form low to high. On the other hand, emotion evaluation shows the emotion color from negative to positive. There are other ways of representing emotions, but besides the classical categorical view, this is mostly used in automatic perception. This is because a few classes of emotion are easier to perceive. Researches in the field of Human-robot interaction have some positive examples of robots that perceive emotions. Here, some of them are presented.

In 2011 the Pet Robot was created. It is a personal robot capable of detection and performing different reactions based on five different emotions: happiness sadness, fear, anger and neutral. The robot reaction is displayed on a LED display. Also, some previously recorded sentences that match the emotions are played. The reported accuracy of this system's emotion classification is 80%.

Another example of practical usage on emotion classifier in robotics is presented by Charles Babbage who developed robotic co-driver and navigator. This robot detects the driver's emotion and reacts accordingly using the following features: tempo, pitch and emphasis and features from human's movement as well. The classifier implemented in this robot works with accuracy of 70%.

In the next section a robotic system built for this research is presented. Even more so, a focus is brought to its evaluation and effect.

IV. INTRODUCING THE EXPERIMENTAL ROBOT

In many cultures, if someone is in need of help, the other gives him/her a hand. This is just one possible human's behaviour that will be used for this research. We created a human-robot interaction application called "Wordless call for help" [15]. A robotic arm, Lynx 5, developed by Lynxmotion is used. It should straighten the hand, go in a handshaking position and put the arm back in the starting position when the speaker has negative emotions expressed by his speech. This actually simulates the giving hand gesture. The automatic emotion evaluation system built in the robotic arm gives the most probable emotion evaluation (positive or negative) from the features extracted. Sound features used by the robotic arm are the one that are found important in some psychological researches for human emotion perception.

After building the robotic arm sensitive to emotion expressed by human's speech, the test phase is conducted. In the test phase a human sound of the interaction with the robot is recorded. In the same time the reaction of the robot is captured. For system estimation, another person evaluates the speaker's emotion simultaneously and without visual contact with the robotic arm. The evaluator gives a signalization of his/hers perception on the evaluation of the speaker's emotion. For each interval with duration of two seconds decisions for the evaluation of the speaker's emotion from the robotic system and the evaluator are taken. Each interval can be evaluated either positive, negative or as a non-voiced interval. With this information gathered in the testing phase for approximately 5 minutes of real-world conversation of the speaker, evaluation of the robotic system is done.

Some statistical measures for measuring the performance of the robot's classifier in relation to the true evaluator's perception are calculated. Positive emotions are classified with greater precision than negative emotions. More than 80% of the intervals that describe positive emotions are captured by the robot. Negative emotions have a recall of 62.2%. The overall accuracy for all non-voiced intervals is 73.14%. Evaluation of similar systems that do not have the property of perceiving spontaneous and natural speech gives accuracy between 75% and 90%. Considering that here the data is real-world, the percentage of the accuracy shows great accomplishment for the automatic emotion perception system used in our application.

V. Experimental Results

More important for robots that can perceive emotion evaluation is the weather their evaluation is similar to human evaluation of the emotion. Indeed, emotion perception is a complex task even for humans. Usually, a human can perceive emotions more precisely on another human if they know or are acquainted to each other. But, no matter how similar emotion expression by humans may be, there are still slight differences between emotion expressions of two different humans. This can be due to different cultural, social and other diversity issues.

In order to explore this problem more deeply, a survey is conducted on the emotion perception system. Indeed, more people were recorded expressing different emotion – positive and negative. Afterwards different human evaluators classified the recorded speech as either positive or negative evaluated emotion. This way the robot evaluator was tested. The idea of the research is to find the correlation between human evaluator's and robot's emotion detection. Experiments are done with the "Wordless call for help" application. The robot's and evaluator's precision are shown on Figure 1. Great recall of 89% is noticed for the positively evaluated emotions. Unfortunately, worse results are gained for the negatively evaluated emotions. This could be because of lack of different negatively evaluated sounds in the training set for the robotic system. Unlike here, in most researches the precision is calculated for only one person. The diversity of speakers used in the testing phase, makes the problem for emotion evaluation classification harder. Because of that the precision presented in Figure 1 is satisfying for the robot. This in general shows some evidence that making emotion aware robots that function like humans is possible and applicable.

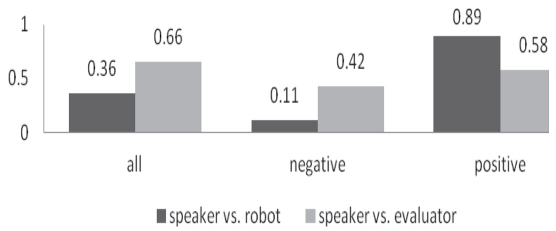

Figure 2: Emotion classification precision of the robot's and evaluator's prediction vs. the speaker's true emotion evaluation statement.

Furthermore, the calculated correlation coefficient, describing the similarity of the robot's classifier and the evaluator's perception on human emotion is 0.62. This shows a good linear relationship and similarity of the perception on emotion evaluation of the robotic system and the human evaluator. Also, the percentage of evaluator's precision is close to the precision of the robotic system evaluator.

VI. Emotion aware robots

The discussion made previously led us to the conclusion that robots can perceive emotions with satisfying precision and as close as humans as possible. We demonstrated one example of a robot application. Robotics is an ongoing field of research and better accomplishment in human-robot interaction are exacted. Here we have straightened the path for building robots that can perceive human emotions. In this section the effects of building such robots are discussed. One question is how people will react on emotion aware robots. From one point of view these robots will find its usage useful in hospitals, nursing facilitates and healthcare institutions as robotic medical assistants. These can alert if a help is needed to the human with which the robot interacts. Other usage is just being a company to these people and controlling their emotional status. The goal is always to bring a positively evaluative emotion to human-robot interaction. From a different perspective, some humans don't want to recognize and appreciate the robots and their intention to improve human life. Not all humans want to have someone to detect their emotions and do something about it. Will this ruin their privacy? Either way, using these robots must always be in service of humans. A policy for usage of emotion information should be taken in consideration.

In the environment, empathic robots can be included in everyday life. In working environments, especially when working with clients and interaction with humans is present, robots can control the communication process. Empathic robots could make sure the clients are well served. In this way the administration can be controlled.

Using robots sensitive to emotions in hospitals for patients that have difficult temper or in psychological institutions is one possible useful application. In order to include robots that perceive emotion in the environment, first introduction to the people must be made. People should know that these robots are present for their service. That is why ethical usage of robots must be obligatory for the robotic industry.

VII. Conclusion

We demonstrated a compassionate human-robot interaction in practice through practical use on emotion perception in robots. Emotion perception of the robot made by human speech that is biologically driven, using the characteristics of the human sound perception, showed great precision of 72%.

The relevance of the robot was evaluated with real-world speech. Also, results showed great similarity of the ability of both robot and human to perceive positive or negative emotions. This confirms the significance of robots and that they can perceive emotions similar to humans. In general, this improves the notion of human-robot interaction.

In the future, emotions differentiation in more categories would be essential for getting more precise emotion perception of the human speech. Also, more research about the emotion perception process could be made in order to increase the precision of the emotion perception systems.

Bringing robots that perceive emotion in the environment should be considered carefully. Rules for their usage must be stated and introduced to humans. This arises many ethical questions that must be considered before globalizing the usage of these smart and empathic robots.

Acknowledgment

This work was partially financed by the Faculty of Computer Science and Engineering at the "Ss.Cyril and Methodius University.

References

[1] V.Kirandziska and N. Ackovska, "Sound features used in emotion classification for human interaction with robotic systems.", in *Proc. Ninth International Conference for Informatics and Information Technologies*, 2012, pp. 91-95.

[2] N. Ackovska and S. Bozinovski, "Biped Robots: From Inverted Pendulum to Programming 12dof Dancing Postures", in *Proc. Seventh International Conference for Informatics and Information Technologies*, 2010, pp.3-7.

[3] T. Vogt, E. Andr´e, and J. Wagner, "*Automatic Recognition of Emotions from Speech: A Review of the Literature and Recommendations for Practical Realization*", Affect and Emotion in HCI. Lecture Notes in Computer Science: Springer-Verlag., Berlin Heidelberg, vol. 4868, pp. 75–91, 2008.

[4] T. Vogt, E. Andre and N. Bee, "EmoVoice – A framework for online recognition of emotions from voice", in *Proc. of the 4th IEEE tutorial and research workshop on Perception and Interactive Technologies for Speech-Based Systems: Perception in Multimodal Dialogue Systems*, 2008, pp. 188 – 199.

[5] F. Hegel, T. Spexard, T. Vogt, G. Horstmann and B. Wrede, "Playing a different imitation game: Interaction with an empathic android robot.", in *Proc. IEEE-RAS International Conference on Humanoid Robots*, 2006, pp. 56-61.

[6] R. W. Picard, "Affective computing", Massachusetts Institute of Technology (MIT), Massachusetts, United States, Media Laboratory Perceptual Computing Section Technical Report No. 321, 1995.

[7] K. R. Scherer, R. Banse, H. G. Wallbott, and T. Goldbeck, " Vocal Cues in Emotion Encoding and Decoding", *Motivation and Emotion*, vol. 15, pp. 123-148 (1991)

[8] K.R. Scherer, "Vocal affect expression: A review and a model for future research". *Psychological Bulletin*, vol. 99, pp.143-165, 1986.

[9] L. Caponetti, C. Buscicchio and G. Castellano, "Biologically inspired emotion recognition from speech", EURASIP journal on Advances in Signal Processing , vol. 2011, 2011.

[10] J. Cichosz and K. Ślot, "Emotion recognition in speech signal using emotion extracting binary decision trees", Institute of Electronics, Technical University of Lodz, Poland, Technical Report, 2010.

[11] K. Dai, H. J. Fell, and J. MacAuslan, "Recognizing Emotion in speech using neural networks", in *Proc. of the IASTED International Conference on Telehealth and Assistive Technologies*, vol. 619, pp. 31-36, 2008.

[12] O. Kwon, K. Chan, J. Hao and T. Lee, "Emotion Recognition by Speech Signals", in *Proc. of Eurospeech, Genewa*, pp. 125-128, 2003.

[13] Y. Huang, G. Zhang, F. Da, "Speech Emotion Recognition with New Spectral Features and Its Application to Pet Robot", *Journal of Computational Information Systems*, vol. 13, pp. 4915-4922, 2011.

[14] E. Mower, M. J. Mataric and S. Narayanan, "A framework for automatic human emotion classification using emotion profiles". *IEEE Transactions on audio, speech and language processing*, vol. 19, no. 5., pp. 1057-1070, July 2011.

[15] V. Kirandziska and N. Ackovska, "Human-robot interaction based on human emotions extracted from speech", in *Proc. Of the 20th Telecommunications Forum - TELFOR,* 2012, In print.

[16] S. Yacoub, S. Simske, X. Lin and J. Burns, "Recognition of Emotions in Interactive Voice Response Systems", HPL Technical Report 136, 2003.